

Heat and Noise on Cubes and Spheres: The Sensitivity of Randomly Rotated Polynomial Threshold Functions

Cristopher Moore
Santa Fe Institute
textttmoore@santafe.edu

Alexander Russell
University of Connecticut
acr@cse.uconn.edu

August 26, 2014

Abstract

We establish a precise relationship between spherical harmonics and Fourier basis functions over a hypercube randomly embedded in the sphere. In particular, we give a bound on the expected Boolean noise sensitivity of a randomly rotated function in terms of its “spherical sensitivity,” which we define according to its evolution under the spherical heat equation. As an application, we prove an average case of the Gotsman-Linial conjecture, bounding the sensitivity of polynomial threshold functions subjected to a random rotation.

1 Introduction

The *average sensitivity* $\text{AS}(f)$ of a Boolean function $f : \mathbb{Z}_2^n \rightarrow \{+1, -1\}$ is n times the probability that $f(x) \neq f(y)$, where x is chosen uniformly at random and y is chosen uniformly from x ’s neighbors at Hamming distance 1. Similarly, the *noise sensitivity* $\text{NS}_\varepsilon(f)$ is the probability that $f(x) \neq f(y)$ where x is uniformly random and y is formed by flipping each bit of x independently with probability ε . Sensitivity is a basic structural characteristic of Boolean functions, with applications to computational complexity, pseudorandomness, machine learning, and the theory of social choice [11, 10, 13, 5, 1, 7, 14].

Many cases of interest focus on threshold functions $f(x)$ defined from a smooth function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ and a threshold value $\theta \in \mathbb{R}$,

$$f(x) = \begin{cases} +1 & \text{if } p(x) \geq \theta, \\ -1 & \text{if } p(x) < \theta. \end{cases}$$

where we define a Boolean function by restricting p , and f , to the hypercube $\{\pm 1\}^n \subset \mathbb{R}^n$. In particular, polynomial threshold functions $f(x) = \text{sgn}(p(x))$, where $p(x)$ is a polynomial of degree d , have played a dominant role in this setting. Gotsman and Linial [4] conjectured that the sensitivity of such functions is maximized when p is a symmetric polynomial whose roots slice the hypercube at d Hamming weights near $n/2$,

$$p(x_1, \dots, x_n) = q\left(\sum_i x_i\right) \quad \text{where} \quad q(x) = \prod_{h=\lfloor (n-d+1)/2 \rfloor}^{\lfloor (n+d-1)/2 \rfloor} \left(x - h - \frac{1}{2}\right).$$

This implies that

$$\text{AS}(f) = O(d\sqrt{n}) \quad \text{and} \quad \text{NS}_\varepsilon(f) = O(d\sqrt{\varepsilon}),$$

where the constant in the O depends neither on d nor n . This is known in the case $d = 1$, i.e., where $f(x)$ is a halfspace [15]. However, for $d > 1$ it has remained open for some time.

The first nontrivial bounds for threshold functions of degree $d > 1$ were obtained quite recently [1, 5], showing

$$\text{AS}(f) = 2^{O(d)} n^{1-\alpha} \log n \quad \text{and} \quad \text{NS}_\varepsilon(f) = 2^{O(d)} \varepsilon^\alpha \log(1/\varepsilon),$$

where $\alpha = O(1/d)$. These bounds work by dividing polynomials into two classes: “juntas” where a few variables are highly influential, and “regular” polynomials where no variable has large influence. The regular case is handled using anticoncentration bounds and the invariance principle of [12], showing that the distribution of values of $p(x)$ is close to what it would be if x were drawn from the Gaussian distribution as opposed to the uniform distribution on the hypercube.

Using different reasoning [9], it was recently shown that

$$\mathbb{AS}(f) = \sqrt{n} (\log n)^{O(d \log d)} 2^{O(d^2 \log d)} \quad \text{and} \quad \mathbb{NS}_\varepsilon(f) = \sqrt{\varepsilon} (\log(1/\varepsilon))^{O(d \log d)} 2^{O(d^2 \log d)}.$$

While this dependence on d is somewhat regrettable, these results show that the Gotsman-Linial conjecture holds, up to polylogarithmic factors, for each fixed d .

Even when our ultimate questions pertain to sensitivity on the hypercube, working with functions defined on \mathbb{R}^n can permit techniques from analysis to be brought to bear on the problem. This has motivated interest in continuous notions of noise sensitivity, most notably *Gaussian sensitivity*, which is obtained by placing a Gaussian measure on \mathbb{R}^n and applying Gaussian noise. In this setting, a simple and elegant argument [8] shows that the Gaussian analog of $\mathbb{NS}_\varepsilon(f)$ is indeed $O(d\sqrt{\varepsilon})$.

In this article, we introduce a notion of *spherical sensitivity* for functions defined on the unit n -sphere. By analyzing how spherical harmonics are carried to Boolean harmonics, i.e., Fourier basis functions over the hypercube \mathbb{Z}_2^n , we give a transfer theorem bounding the Boolean sensitivity in terms of the spherical sensitivity. Our results hold in expectation, when the function, or equivalently the hypercube, is randomly rotated in \mathbb{R}^n . In essence, we show that the distribution of angles induced by Boolean noise on the hypercube can be modeled by the effect of Brownian motion on the sphere, or equivalently diffusion driven by the spherical heat equation.

As an application, by bounding the spherical sensitivity of polynomial threshold functions, we establish the Gotsman-Linial conjecture on average, in the following sense: for any polynomial p of degree d , if we apply a random rotation R and then restrict to the hypercube, the expected average sensitivity and noise sensitivity of the resulting Boolean threshold function $f(x) = \text{sgn}(Rp(x))$ are $\mathbb{E}_R \mathbb{AS}(f) = O(d\sqrt{n})$ and $\mathbb{E}_R \mathbb{NS}_\varepsilon = O(\varepsilon\sqrt{n})$ respectively.

2 Spherical harmonics and the heat equation

Here and in later sections we repeat information from classic texts [18, 19] and two excellent reviews [2, 3]. Recall that the Laplace operator on \mathbb{R}^n is defined as

$$\Delta_{\mathbb{R}^n} = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}.$$

The Laplace-Beltrami operator on S_{n-1} consists of the contribution to $\Delta_{\mathbb{R}^n}$ arising from the dependence of a function on angular variables rather than the distance from the origin. It can be defined by writing $\Delta_{\mathbb{R}^n}$ in polar coordinates,

$$\Delta_{\mathbb{R}^n} = \frac{\partial}{\partial r^2} + \frac{n-1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \Delta_{S_{n-1}}.$$

A function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is *harmonic* if $\Delta_{\mathbb{R}^n} h = 0$. In particular, for each $\ell \geq 0$, there is a linear subspace of harmonic homogeneous polynomials of degree ℓ of dimension d_ℓ (where we assume $n \geq 3$)

$$d_\ell = \binom{n+\ell-1}{\ell} - \binom{n+\ell-3}{\ell-2} = \frac{n+2\ell-2}{n-2} \binom{n+\ell-3}{\ell}. \quad (1)$$

Restricting these polynomials to S_{n-1} gives the so-called *spherical harmonics*. We denote a basis for these as $\{Y_{\ell,j} \mid 1 \leq j \leq d_\ell\}$. They are eigenfunctions of the Laplace-Beltrami operator:

$$\Delta_{S_{n-1}} Y_{\ell,j} = -\ell(n+\ell-2) Y_{\ell,j}. \quad (2)$$

Any function g in $L_2(S_{n-1})$ can be expanded in terms of spherical harmonics,

$$g = \sum_{\ell \geq 0} g_\ell = \sum_{\ell \geq 0} \sum_{j=1}^{d_\ell} \hat{g}(\ell, j) Y_{\ell,j}. \quad (3)$$

This is analogous to the expansion into Fourier series over $[0, 1) \subset \mathbb{R}$, where small ℓ corresponds to smooth, low-frequency variations, and larger ℓ corresponds to higher frequencies.

3 Noise sensitivity and the heat equation on the cube and the sphere

The standard notion of noise sensitivity for a Boolean function $f : \mathbb{Z}_2^n \rightarrow \{\pm 1\}$ is as follows [14]. Define a linear operator K_ε on the space of probability distributions over \mathbb{Z}_2^n that independently flips each bit with probability ε . That is, if $d(x, y)$ denotes the Hamming distance,

$$K_\varepsilon(x, y) = \varepsilon^{d(x, y)} (1 - \varepsilon)^{n - d(x, y)}.$$

Then if δ_x is the Kronecker delta function where $\delta_x(z) = 1$ if $z = x$ and 0 otherwise,

$$\text{NS}_\varepsilon(f) = \mathbb{E}_{x, y} \Pr[f(y) \neq f(x)],$$

where x is uniform in $\{\pm 1\}^n$ and y is chosen from the distribution $K_\varepsilon \delta_x$. Equivalently, if we define the inner product on the cube as

$$\langle f, g \rangle = \mathbb{E}_{x \in \mathbb{Z}_2^n} f(x)^* g(x),$$

then

$$\text{NS}_\varepsilon(f) = \frac{1}{2} (1 - \langle f, K_\varepsilon f \rangle).$$

We write f in the Fourier basis, expanding in terms of the characters χ_k ,

$$f(x) = \sum_{k \in \mathbb{Z}_2^n} \widehat{f}(k) \chi_k(x) \quad \text{where} \quad \chi_k(x) = (-1)^{k \cdot x} \quad \text{and} \quad \widehat{f}(k) = \langle f, \chi_k \rangle.$$

Then we can use the fact that χ_k is an eigenvector of K_ε ,

$$K_\varepsilon \chi_k = (1 - 2\varepsilon)^{|k|} \chi_k, \tag{4}$$

where $|k|$ denotes the Hamming weight of the frequency vector k . Then we obtain

$$\begin{aligned} \text{NS}_\varepsilon(f) &= \frac{1}{2} \left(1 - \sum_{k \in \mathbb{Z}_2^n} |\widehat{f}(k)|^2 (1 - 2\varepsilon)^{|k|} \right) \\ &= \frac{1}{2} \sum_{k \in \mathbb{Z}_2^n} |\widehat{f}(k)|^2 (1 - (1 - 2\varepsilon)^{|k|}). \end{aligned} \tag{5}$$

Here we used the fact that $\langle f, f \rangle = \sum_k |\widehat{f}(k)|^2 = 1$ since $|f(x)|^2 = 1$ for all x . However, we can also take (5) as the definition of $\text{NS}_\varepsilon(f)$, in which case it can be applied to any function $f : \mathbb{Z}_2^n \rightarrow \mathbb{C}$.

We can also write the noise sensitivity in terms of a continuous-time (but discrete-space) heat equation on the hypercube. Let A be the adjacency matrix of the hypercube, and let

$$L = A - n\mathbf{1}$$

be the graph Laplacian. If we apply the heat equation

$$\frac{\partial f}{\partial t} = Lf$$

for time ε' with initial condition $f(0) = f$, we have

$$f(\varepsilon') = e^{\varepsilon' L} f.$$

The characters χ_k are eigenfunctions of the Laplacian,

$$L\chi_k = -2|k|\chi_k \quad \text{and} \quad e^{\varepsilon' L} \chi_k = e^{-2|k|\varepsilon'} \chi_k.$$

Matching these eigenvalues with (4), we see that if

$$\varepsilon' = \frac{1}{2} \log \frac{1}{1 - 2\varepsilon} = \varepsilon + O(\varepsilon^2), \tag{6}$$

then

$$K_\varepsilon = e^{\varepsilon' L},$$

and

$$\begin{aligned} \mathbb{NS}_\varepsilon &= \frac{1}{2} (1 - \langle f, f(\varepsilon') \rangle) \\ &= \frac{1}{2} (1 - \langle f, e^{\varepsilon' L} f \rangle) \\ &= \frac{1}{2} \left(1 - \sum_{k \in \mathbb{Z}_2^n} |\widehat{f}(k)|^2 e^{-2\varepsilon' |k|} \right) \\ &= \frac{1}{2} \sum_{k \in \mathbb{Z}_2^n} |\widehat{f}(k)|^2 (1 - e^{-2\varepsilon' |k|}). \end{aligned} \tag{7}$$

In analogy with this heat-equation picture of the Boolean noise sensitivity, we define the *spherical sensitivity* $\mathbb{SS}_t(g)$ of a function $g : S_{n-1} \rightarrow \{\pm 1\}$ as follows. First define the inner product of two functions $f, g : S_{n-1} \rightarrow \mathbb{C}$ as

$$\langle f, g \rangle_S = \mathbb{E}_{x \in S_{n-1}} f(x)^* g(x) = \frac{1}{\Omega_{n-1}} \int_{S_{n-1}} dx f(x)^* g(x), \tag{8}$$

where Ω_{n-1} denotes the surface area of S_{n-1} ,

$$\Omega_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}. \tag{9}$$

The heat equation on S_{n-1} is

$$\frac{\partial g}{\partial t} = \Delta_{S_{n-1}} g,$$

and applying it for time t with the initial condition $g(0) = g$ gives

$$g(t) = e^{\Delta_{S_{n-1}} t} g.$$

Then we define $\mathbb{SS}_t(g)$ as

$$\begin{aligned} \mathbb{SS}_t(g) &= \frac{1}{2} (1 - \langle g, g(t) \rangle_S) \\ &= \frac{1}{2} (1 - \langle g, e^{t \Delta_{S_{n-1}}} g \rangle_S) \\ &= \frac{1}{2} \sum_{\ell, j} |\widehat{g}(\ell, j)|^2 (1 - e^{-t \ell(n+\ell-2)}), \end{aligned} \tag{10}$$

where we used (2) and the expansion (3).

Here we assumed that g takes values in $\{\pm 1\}$. In that case, $\mathbb{SS}_t(g)$ is the probability that $g(x) \neq g(y)$ if x is uniformly random and y is the position of a particle that starts at x and undergoes Brownian motion for time t . However, as with (5), we will take (10) as the definition of $\mathbb{SS}_t(g)$, thus extending the notion of sensitivity to arbitrary functions $g : S_{n-1} \rightarrow \mathbb{C}$.

Comparing (7) and (10), we see that flipping bits with probability ε is roughly analogous to running the heat equation on the sphere for time $t = O(\varepsilon/n)$. We will tighten this analogy in Theorem 1 below.

4 Zonal harmonics and Gegenbauer polynomials

In this section and the next, we continue our review of spherical harmonics and their associated orthogonal polynomials [18, 19, 2, 3]. We include somewhat more machinery than is strictly necessary to prove our main result. However, in many cases this machinery gives us a more explicit picture of what is going on, and may be useful in proving more detailed results.

Let $\eta = (1, 0, \dots, 0)$ be the north pole. If $f(Rx) = f(x)$ for all rotation matrices that fix η , then $f(x)$ depends only on $z = \eta \cdot x$. Such functions are called *zonal*. The inner product of two such functions can be written as a weighted inner product over the interval $-1 \leq z \leq 1$,

$$\langle f, g \rangle_S = \frac{\Omega_{n-2}}{\Omega_{n-1}} \int_{-1}^1 dz w_\alpha(z) f(z)^* g(z), \quad (11)$$

where we wantonly abuse notation by identifying $f(z)$ with $f(x)$, and the weight

$$w_\alpha(z) = (1 - z^2)^{\alpha - \frac{1}{2}} \quad (12)$$

and the constant Ω_{n-2} account for the volume of the annulus between height z and $z + dz$.

There is a unique harmonic polynomial of each degree ℓ . A nice orthogonal family of such polynomials, called the *zonal spherical harmonics* or the *ultraspherical* or *Gegenbauer* polynomials, are as follows. For historical reasons, we parametrize them with a half-integer α rather than the integer dimension n :

$$\alpha = \frac{n}{2} - 1,$$

in which case the dimension (1) becomes

$$d_\ell = \frac{\alpha + \ell}{\alpha} \binom{2\alpha + \ell - 1}{\ell}. \quad (13)$$

For each value of α , we have a family of polynomials $\{\gamma_\ell \mid \ell \in \mathbb{N}\}$ of degree ℓ ,

$$\gamma_\ell(z) = \frac{1}{\sqrt{N^{(\ell)}}} G^{(\ell)}(z) \quad \text{where} \quad G^{(\ell)}(z) = \sum_{k=0}^{\lfloor \ell/2 \rfloor} (-1)^k \frac{(\ell - k + \alpha - 1)!}{(\alpha - 1)! k! (\ell - 2k)!} (2z)^{\ell - 2k} \quad (14)$$

$$\begin{aligned} \text{and} \quad N^{(\ell)} &= \langle G^{(\ell)}, G^{(\ell)} \rangle \\ &= \frac{\Omega_{n-2}}{\Omega_{n-1}} \times \frac{\pi 2^{1-2\alpha} \Gamma(\ell + 2\alpha)}{(\ell + \alpha) \Gamma(\ell + 1) \Gamma(\alpha)^2} \\ &= \frac{\alpha}{\alpha + \ell} \binom{2\alpha + \ell - 1}{\ell}. \end{aligned} \quad (15)$$

Note that $G^{(\ell)}$ is the ℓ th Legendre polynomial when $\alpha = 1/2$ (i.e., when $n = 3$). Then for each α , the γ_ℓ are orthonormal with respect to the inner product (11):

$$\langle \gamma_\ell, \gamma^{(m)} \rangle_S = \delta_{\ell, m}.$$

Thus, given a zonal function $f(z)$, we can write

$$f(z) = \sum_{\ell \geq 0} \hat{f}^{(\ell)} \gamma_\ell(z) \quad \text{where} \quad \hat{f}^{(\ell)} = \langle f, \gamma_\ell \rangle_S.$$

This transform is unitary, so inner products are preserved:

$$\langle f, g \rangle_S = \sum_{\ell \geq 0} \hat{f}^{(\ell)*} \hat{g}^{(\ell)}.$$

The Gegenbauer polynomials have deep roots in the representation theory of Lie groups. Let SO_n be the group of orthogonal rotations of \mathbb{R}^n ; then the harmonic polynomials of degree ℓ form an irreducible representation ρ_ℓ of SO_n with dimension d_ℓ . We can think of $L_2(S_{n-1})$ as the subspace of $L_2(\text{SO}_n)$ consisting of functions that are right-invariant under the subgroup SO_{n-1} that fixes the north pole η ; that is, functions $f(R)$ that only depend on $R\eta$. The Gegenbauer polynomials span the subspace of $L_2(\text{SO}_n)$ consisting of functions which are left- and right-invariant under SO_{n-1} ; that is, functions that are zonal, depending only on $z = \langle \eta \cdot R\eta \rangle$, or equivalently on the latitude of $R\eta$. The fact that γ_ℓ is the unique zonal polynomial of degree ℓ corresponds to the fact that SO_n and SO_{n-1} form a Gel'fand pair, i.e., this subspace is one-dimensional.

5 Schur's lemma and evaluation maps

Of course, η is an arbitrary choice for the north pole. For any $w \in S_{n-1}$, there is a unique polynomial of degree ℓ which is zonal around w , i.e., which is fixed under the copy of SO_{n-1} that preserves w . Given a function f and $R \in \text{SO}_n$, define Rf as the function

$$Rf(x) = f(R^{-1}x).$$

Now let $R \in \text{SO}_n$ be any rotation such that $R\eta = w$. Then if we write $\gamma_\ell^{(\eta)}(x) = \gamma_\ell(\eta \cdot x)$, we can define $\gamma_\ell^{(w)} = R\gamma_\ell^{(\eta)}$, so that

$$\gamma_\ell^{(w)}(x) = \gamma_\ell^{(\eta)}(R^{-1}x) = \gamma_\ell(w \cdot x).$$

The inner products of these functions are again given by a Gegenbauer polynomial: for any $w, y \in S_{n-1}$,

$$\left\langle \gamma_\ell^{(w)}, \gamma_\ell^{(y)} \right\rangle_S = \frac{1}{\Omega_{n-1}} \int_{S_{n-1}} dx \gamma_\ell(w \cdot x) \gamma_\ell(y \cdot x) = \frac{1}{\sqrt{d_\ell}} \gamma_\ell(w \cdot y). \quad (16)$$

Taking $w = y$ tells us how large γ_ℓ gets at the poles:

$$\gamma_\ell(1) = |\gamma_\ell(-1)| = \sqrt{d_\ell}. \quad (17)$$

In addition, if we fix w and take the expectation over y of the inner product squared, we get

$$\mathbb{E}_y \left| \left\langle \gamma_\ell^{(w)}, \gamma_\ell^{(y)} \right\rangle_S \right|^2 = \frac{1}{d_\ell} \left| \langle \gamma_\ell, \gamma_\ell \rangle_S \right|^2 = \frac{1}{d_\ell}. \quad (18)$$

Since averaging over y is the same as averaging over R according to the Haar measure on SO_n , this is equivalent to

$$\mathbb{E}_R \left| \langle \gamma_\ell, R\gamma_\ell \rangle_S \right|^2 = \frac{1}{d_\ell}. \quad (19)$$

This holds more generally for any spherical harmonic f_ℓ of degree ℓ ,

$$\mathbb{E}_R \left| \langle f_\ell, Rf_\ell \rangle_S \right|^2 = \frac{\|f_\ell\|_S^2}{d_\ell}. \quad (20)$$

where $\|f_\ell\|_S^2 = \langle f_\ell, f_\ell \rangle_S = \mathbb{E}_{x \in S_{n-1}} |f_\ell(x)|^2$. This is a form of Schur's lemma: for any vector v belonging to an irreducible representation ρ of a group G , we have $\mathbb{E}_g |\langle v, gv \rangle|^2 = |v|^2/d_\rho$.

More generally, let M be a linear operator on $L_2(S_{n-1})$. Conjugating it with a random rotation yields an operator which commutes with all rotations. By Schur's lemma any such operator is block diagonal, where each block is a scalar matrix operating on the degree- ℓ spherical harmonics. Thus

$$\mathbb{E}_R R^{-1} M R = \bigoplus_{\ell \geq 0} M_\ell \quad \text{where} \quad M_\ell = \frac{\text{tr } M_\ell}{d_\ell} \mathbb{1}_\ell,$$

where $\mathbb{1}_\ell$ is the projection operator onto the space of degree- ℓ harmonics. Thus if $f = \sum_{\ell \geq 0} f_\ell$ where each f_ℓ is a spherical harmonic of degree ℓ ,

$$\begin{aligned} \mathbb{E}_R \langle Rf, M \cdot Rf \rangle_S &= \left\langle f, \left(\mathbb{E}_R R^{-1} M R \right) f \right\rangle_S \\ &= \sum_{\ell \geq 0} \langle f_\ell, M_\ell f_\ell \rangle_S \\ &= \sum_{\ell \geq 0} \mathbb{E}_R \langle Rf_\ell, M \cdot Rf_\ell \rangle_S. \end{aligned} \quad (21)$$

In particular, if $f = \sum_{\ell} f_\ell$ then the expected outer product of Rf with itself is

$$\mathbb{E}_R |Rf\rangle \langle Rf| = \sum_{\ell \geq 0} \frac{\|f_\ell\|_S^2}{d_\ell} \mathbb{1}_\ell. \quad (22)$$

Thus if $g = \sum_{\ell} g_{\ell}$ and $h = \sum_{\ell} h_{\ell}$,

$$\mathbb{E}_R \langle g, Rf \rangle_S \langle Rf, h \rangle_S = \sum_{\ell \geq 0} \frac{\|f\|_S^2}{d_{\ell}} \langle g_{\ell}, h_{\ell} \rangle_S. \quad (23)$$

Another consequence of the irreducibility of ρ_{ℓ} is that any linear operator from ρ_{ℓ} to \mathbb{C} can be written as $\langle \phi, \cdot \rangle_S$ where $\phi = \sum_{y \in S_{n-1}} c_y \gamma_{\ell}^{(y)}$ for some finite set of nonzero coefficients c_y . In particular, for any $y \in S_{n-1}$ and any $\ell \geq 0$ there is an *evaluation map* $\mathcal{E}_{\ell}^{(y)} : \rho_{\ell} \rightarrow \mathbb{C}$ such that, for any spherical harmonic f of degree ℓ , we have $\mathcal{E}_{\ell}^{(y)}(f) = f(y)$. We can express it as

$$\mathcal{E}_{\ell}^{(y)}(f) = f(y) = \sqrt{d_{\ell}} \left\langle \gamma_{\ell}^{(y)}, f \right\rangle_S. \quad (24)$$

To see this, think of y as the north pole, and consider an orthonormal basis that includes $\gamma_{\ell}^{(y)}$. Since $\gamma_{\ell}^{(y)}$ is the unique harmonic of degree ℓ that is zonal around y , all other basis functions are zero at y ; otherwise they would have a nonzero projection onto $\gamma_{\ell}^{(y)}$ if we average over the subgroup SO_{n-1} of rotations that preserve y . The normalization $\sqrt{d_{\ell}}$ then follows from (17).

By summing over all ℓ , we can similarly express the evaluation map for all integrable functions $f \in L_2(S_{n-1})$. That is, we can define the evaluation map

$$\mathcal{E}^{(y)} = \sum_{\ell \geq 0} \mathcal{E}_{\ell}^{(y)}.$$

Then

$$\mathcal{E}^{(y)}(f) = f(y) = \sum_{\ell \geq 0} \sqrt{d_{\ell}} \left\langle \gamma_{\ell}^{(y)}, f \right\rangle_S. \quad (25)$$

To put it differently, we can express the Dirac delta function as

$$\delta(x - y) = \sum_{\ell \geq 0} \sqrt{d_{\ell}} \gamma_{\ell}^{(y)}(x) = \sum_{\ell \geq 0} \sqrt{d_{\ell}} \gamma_{\ell}(y \cdot x). \quad (26)$$

6 Relating noise sensitivity and spherical sensitivity for randomly rotated functions

In this section we will prove our main transfer theorem, bounding the expected noise sensitivity of a randomly rotated function in terms of its spherical sensitivity.

We identify the hypercube \mathbb{Z}_2^n with the set $H = \{\pm 1/\sqrt{n}\}^n$ lying on the unit sphere. If we restrict the inner product to this set, we obtain a cubical inner product, which we write

$$\langle f, g \rangle_H = \mathbb{E}_{x \in H} f(x)^* g(x).$$

In particular, for each frequency vector $k \in \mathbb{Z}_2^n$, if we extend the character χ_k to the sphere as a multilinear function of degree $|k|$,

$$\chi_k(x) = \prod_{i: k_i=1} \sqrt{n} x_i,$$

then the χ_k are orthonormal with respect to the cubical inner product,

$$\langle \chi_k, \chi_{k'} \rangle_H = \delta_{k, k'}.$$

If we define a Boolean function $f|_H$ by restricting a function f to the hypercube, its Fourier coefficients are

$$\widehat{f}|_H(k) = \langle f, \chi_k \rangle_H.$$

The *energy* of f at the character k is $|\widehat{f}(k)|^2$.

In order to bound the noise sensitivity, we need to compute the expected energy of Rf where R is uniformly random, i.e., chosen according to the Haar measure in SO_n ; or equivalently, the expected

energy of f 's restriction to a randomly rotated hypercube. One simple observation is the following. Since uniformly rotating any point on the cube yields a uniformly random point on the sphere, cubical inner products are equal to spherical inner products in expectation,

$$\mathbb{E}_R \langle Rf, Rg \rangle_H = \langle f, g \rangle_S. \quad (27)$$

In Appendix A we give a precise expression for the expected energy of a randomly rotated spherical harmonic. However, here we just need a few facts. First, if we decompose a function into spherical harmonics, then its expected energy at each k is the sum of the expected energies of its harmonics:

Lemma 1. Let $f \in L_2(S_{n-1})$, and write $f = \sum_{\ell \geq 0} f_\ell$ where each f_ℓ is a spherical harmonic of degree ℓ . Let $R \in \text{SO}_n$ be uniform in the Haar measure. Then for any $k \in \mathbb{Z}_2^n$,

$$\mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 = \sum_{\ell \geq 0} \mathbb{E}_R \left| \widehat{Rf_\ell}|_H(k) \right|^2.$$

Proof. First we use the evaluation maps of Section 5 to write $\widehat{f}(k)$ as a spherical inner product rather than a cubical one. Using (25) we have

$$\widehat{f}(k) = \langle \chi_k, f \rangle_H = \mathbb{E}_{x \in H} \chi_k(x) \mathcal{E}^{(x)}(f) = \langle \psi_k, f \rangle_S,$$

where

$$\psi_k = \mathbb{E}_{x \in H} \chi_k(x) \sum_{\ell \geq 0} \sqrt{d_\ell} \gamma_\ell^{(x)}.$$

Then applying Schur's lemma (21) to the linear operator $M = |\psi_k\rangle \langle \psi_k|$ gives

$$\begin{aligned} \mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 &= \mathbb{E}_R \langle Rf, \psi_k \rangle_S \langle \psi_k, Rf \rangle_S \\ &= \sum_{\ell \geq 0} \mathbb{E}_R \langle Rf_\ell, \psi_k \rangle_S \langle \psi_k, Rf_\ell \rangle_S \\ &= \sum_{\ell \geq 0} \mathbb{E}_R \left| \widehat{Rf_\ell}|_H(k) \right|^2, \end{aligned}$$

completing the proof. \square

Secondly, restricting a spherical harmonic of degree ℓ to the hypercube can only give it nonzero energy at characters of Hamming weight ℓ or less:

Lemma 2. Let f be a spherical harmonic of degree ℓ . If $|k| \geq \ell$, then $\widehat{f}|_H(k) = 0$.

Proof. Restricting any polynomial f of degree ℓ to $H = \{\pm 1/\sqrt{n}\}^n$ imposes the relations $x_i^2 = 1/n$ for all i , so there is a multilinear polynomial f' of degree at most ℓ such that $f|_H = f'|_H$. Each multilinear monomial is proportional to a character χ_k with $|k| \leq \ell$, and these are orthogonal to all χ_k with $|k| > \ell$. \square

Thirdly, for any integrable function f , the expected energy of Rf summed over all characters of \mathbb{Z}_2^n equals its norm on the sphere:

Lemma 3. Let $f \in L_2(S_{n-1})$, and let $R \in \text{SO}_n$ be uniform in the Haar measure. Then

$$\mathbb{E}_R \sum_{k \in \mathbb{Z}_2^n} \left| \widehat{Rf}|_H(k) \right|^2 = \|f\|_S^2.$$

Proof. Since the χ_k are orthonormal with respect to the inner product on the hypercube, for any f we have

$$\sum_{k \in \mathbb{Z}_2^n} \left| \widehat{f}|_H(k) \right|^2 = \sum_k \langle f, \chi_k \rangle_H \langle \chi_k, f \rangle_H = \langle f, f \rangle_H$$

Then (27) gives

$$\mathbb{E}_R \sum_{k \in \mathbb{Z}_2^n} \left| \widehat{Rf}|_H(k) \right|^2 = \mathbb{E}_R \langle Rf, Rf \rangle_H = \langle f, f \rangle_S = \|f\|_S^2. \quad \square$$

We are now ready to prove our transfer theorem, which bounds the expected noise sensitivity and average sensitivity in terms of the spherical sensitivity.

Theorem 1. Let $f \in L_2(S_{n-1})$, and let $R \in \text{SO}_n$ be uniform in the Haar measure. Then

$$\mathbb{E}_R \text{NS}_\varepsilon(Rf|_H) \leq \text{SS}_t(f), \quad (28)$$

where

$$t = \frac{1}{n} \log \frac{1}{1-2\varepsilon} = \frac{2\varepsilon}{n} (1 + O(\varepsilon)). \quad (29)$$

Proof. Starting with the Fourier-theoretic expression for the noise sensitivity (5), we have

$$\begin{aligned} \mathbb{E}_R \text{NS}_\varepsilon(Rf|_H) &= \frac{1}{2} \mathbb{E}_R \sum_{k \in \mathbb{Z}_2^n} \left| \widehat{Rf|_H}(\chi) \right|^2 (1 - (1-2\varepsilon)^{|k|}) \\ &= \frac{1}{2} \mathbb{E}_R \sum_k \sum_{\ell \geq 0} \left| \widehat{Rf_\ell|_H}(k) \right|^2 (1 - (1-2\varepsilon)^{|k|}) \end{aligned} \quad (30)$$

$$= \frac{1}{2} \mathbb{E}_R \sum_\ell \sum_{k: |k| \leq \ell} \left| \widehat{Rf_\ell|_H}(k) \right|^2 (1 - (1-2\varepsilon)^{|k|}) \quad (31)$$

$$\begin{aligned} &\leq \frac{1}{2} \mathbb{E}_R \sum_\ell \sum_k \left| \widehat{Rf_\ell|_H}(k) \right|^2 (1 - (1-2\varepsilon)^\ell) \\ &\leq \frac{1}{2} \sum_\ell \|f_\ell\|_S^2 (1 - (1-2\varepsilon)^\ell). \end{aligned} \quad (32)$$

Here we used Lemma 1 in (30), Lemma 2 in (31), and Lemma 3 in (32).

On the other hand, using $\|f_\ell\|_S^2 = \sum_j |\widehat{f}(\ell, j)|^2$ in (10) gives

$$\begin{aligned} \text{SS}_t(f) &= \frac{1}{2} \sum_\ell \|f_\ell\|_S^2 (1 - e^{-t\ell(n+\ell-2)}) \\ &\geq \frac{1}{2} \sum_\ell \|f_\ell\|_S^2 (1 - e^{-t\ell n}). \end{aligned} \quad (33)$$

Setting $1 - 2\varepsilon = e^{-tn}$ completes the proof of (28). \square

Although we don't need it below, we record an analogous theorem regarding the expected average sensitivity.

Theorem 2. Let $f \in L_2(S_{n-1})$, and let $R \in \text{SO}_n$ be uniform in the Haar measure. Then for any $\alpha > 0$,

$$\mathbb{E}_R \text{AS}(Rf|_H) \leq \frac{2n}{1 - e^{-\alpha}} \text{SS}_{\alpha/n^2}(f). \quad (34)$$

Proof. Recall the Fourier-theoretic expression for the average sensitivity,

$$\begin{aligned} \text{AS}(f) &= -\frac{1}{2} \langle f, Lf \rangle \\ &= \sum_{k \in \mathbb{Z}_2^n} |\widehat{f}(k)|^2 |k|. \end{aligned} \quad (35)$$

Applying the same lemmas to (35) as we did to (5) in the proof of the previous lemma, and noting that $|k| \leq n$, gives

$$\mathbb{E}_R \text{AS}(Rf|_H) \leq \sum_\ell \|f_\ell\|_S^2 \max(\ell, n).$$

Setting $t = \alpha/n^2$ in (33), we have for all $0 \leq \ell \leq n$

$$\frac{1 - e^{-\alpha\ell/n}}{1 - e^{-\alpha}} \geq \frac{\ell}{n},$$

since $1 - e^{-x}$ is concave. This completes the proof of (34). \square

7 Application: the Gotsman-Linial conjecture on average

In this section we bound the spherical sensitivity of polynomial threshold functions, and apply Theorem 1 to bound their expected noise sensitivity.

Our strategy for bounding the spherical sensitivity is similar to that of Kane [8], who proved a similar bound on their Gaussian sensitivity. He used the fact that adding Gaussian noise to a point u can be thought of as choosing a random line through u , and then moving along that line by a distance r chosen from the chi-squared distribution: then $p(x)$ can only change sign if it has a root on the intervening line segment. Similarly, we add noise on the sphere by choosing a random great circle that passes through u , and then moving an angle r along that circle where r is chosen from a distribution derived from the heat equation on S_{n-1} . The sensitivity is then at most the probability that the polynomial has a root on the resulting segment of the great circle.

First we need to bound the distribution of the angle, or equivalently the geodesic distance, that we travel on the sphere in time t . Given two points $u, v \in S_{n-1}$ and a time $t \geq 0$, let $K_t(u, v)$ denote the *heat kernel* on S_{n-1} after time t . That is, for any fixed t and u , $K_t(u, v)$ is the probability distribution of the position v of a particle that starts at u and undergoes Brownian motion for time t . As a linear operator, it is the solution to the partial differential equation

$$\frac{\partial K_t}{\partial t} = \Delta_{S_{n-1}} K_t$$

with the initial condition $K_0 = \mathbb{1}$, so

$$K_t(u, v) = e^{t\Delta_{S_{n-1}}}.$$

By comparing the heat kernel on S_{n-1} to that on \mathbb{R}^{n-1} , we prove the following:

Lemma 4. Fix $u \in S_{n-1}$, and suppose that $v \in S_{n-1}$ is chosen with probability distribution $K_t(u, v)$. Let r denote the angle between u and v . Then

$$\mathbb{E}[r] \leq \sqrt{2(n-1)t}.$$

Proof. In Appendix B we offer an elementary calculus proof that $\mathbb{E}[r^2] \leq 2(n-1)t$, which implies the lemma. However, we can prove something much stronger: namely, that r is stochastically dominated by the corresponding process on the flat tangent space \mathbb{R}^{n-1} . The heat equation on \mathbb{R}^{n-1} is driven by the Laplacian

$$\Delta_{\mathbb{R}^{n-1}} = \sum_{i=1}^{n-1} \frac{\partial^2}{\partial x_i^2}. \quad (36)$$

Place u at the origin, and let r denote the distance from the origin. Since $K_t(u, v) = f(r)$ is spherically symmetric, transforming to polar coordinates gives

$$\frac{\partial f}{\partial t} = \Delta_{\mathbb{R}^{n-1}} f = \frac{\partial^2 f}{\partial r^2} + (n-2) \frac{1}{r} \frac{\partial f}{\partial r}.$$

Similarly, for S_{n-1} , place u at the north pole and let r denote the angle between u and v . Then $K_t(u, v) = f(r)$ is a zonally symmetric function, and applying the Laplace-Beltrami operator gives

$$\frac{\partial f}{\partial t} = \Delta_{S_{n-1}} f = \frac{\partial^2 f}{\partial r^2} + (n-2) \frac{\cos r}{\sin r} \frac{\partial f}{\partial r}. \quad (37)$$

We can view (36) and (37) as governing the probability distributions of two stochastic processes on r . These are well known in the theory of Brownian motion, and are referred to as Bessel and Jacobi processes respectively. Since $\cos r / \sin r \leq 1/r$, the comparison theorem of stochastic differential equations [6] implies that the distribution of r on S_{n-1} is stochastically dominated by its distribution on \mathbb{R}^{n-1} . In particular, its second moment is at most the variance of $n-1$ independent variables $x_1, \dots, x_{n-1} \in \mathbb{R}$ of variance $2t$, giving

$$\mathbb{E}[r^2] \leq 2(n-1)t.$$

Noting that $\mathbb{E}[r] \leq \sqrt{\mathbb{E}[r^2]}$ completes the proof. \square

We remark that for small n , we get a small improvement by computing $\mathbb{E}[r]$ exactly on \mathbb{R}^{n-1} . The fact that $r^2/2$ follows a chi-squared distribution with $n-1$ degrees of freedom implies

$$\mathbb{E}[r] \leq \frac{2\Gamma(n/2)}{\Gamma((n-1)/2)} \sqrt{t} = (1 - O(1/n)) \sqrt{2(n-1)t}.$$

Lemma 5. Let $p \in \mathbb{R}[x_1, \dots, x_n]$ be a polynomial of degree d , and let G a great circle on S_{n-1} . If p is not identically zero on G , then p has no more than $2d$ roots on G .

Proof. Since applying a linear transformation to x_1, \dots, x_n doesn't change p 's degree, without loss of generality we can assume that G is the unit circle in the plane spanned by the x_1 and x_2 axes: that is, the variety, or set of roots, of the polynomial

$$q(x_1, x_2) = x_1^2 + x_2^2 - 1.$$

Restricting to this plane, i.e., setting $x_i = 0$ for all $i > 2$, yields a polynomial $r(x_1, x_2)$ of degree $d_r \leq d$. Bézout's theorem [17, 16] states that two polynomials q, r of degree d_q and d_r can share at most $d_q d_r$ roots unless they share a common factor. It is easy to check that q is irreducible over \mathbb{R} (and even over \mathbb{C}), since if it had a linear factor then G would consist of the union of two lines. Therefore, q and r share a common factor only if q divides r , in which case p is identically zero on G . If they do not, they share at most $2d$ roots. \square

Putting these lemmas together gives us a bound on the spherical sensitivity of a polynomial threshold function.

Theorem 3. Let $p : \mathbb{R}^n \rightarrow \mathbb{R}$ be a polynomial of degree d and let $f(x) = \text{sgn}(p(x))$ where $\text{sgn } z = +1$ for $z \geq 0$ and $\text{sgn } z = -1$ for $z < 0$. Then

$$\text{SS}_t(f) \leq \frac{d}{\pi} \sqrt{2nt}.$$

Proof. Recall that $\text{SS}_t(f)$ is the probability that $f(u) \neq f(v)$ where u is chosen uniformly and v is chosen from $K_t(u, v)$. Equivalently, we can choose u uniformly, and then arrive at v by choosing a uniformly random great circle G passing through u (by choosing a tangent vector from the Haar measure on S_{n-2}), choosing r according to the heat kernel, and moving an angle r along G .

If p is identically zero on S_{n-1} , then f is identically zero as well, in which case $\text{SS}_t(f) = 0$. Otherwise, with probability 1 we have $f(u) \neq 0$, in which case p is not identically zero on G . By Lemma 5, there are at most $2d$ roots of p on G . The probability that $p(u)$ and $p(v)$ have different signs is then at most the expected number of roots of p on G between u and v . Since u 's position on G is uniformly random, this is simply $(2d/2\pi) \mathbb{E}[r]$, which by Lemma 4 is at most $(d/\pi) \sqrt{2nt}$. \square

Theorems 1 and 3 immediately imply our bound on the expected noise sensitivity:

Theorem 4. Let p be a polynomial of degree d and let $f(x) = \text{sgn}(p(x))$. Let $R \in \text{SO}_n$ be uniform in the Haar measure. Then

$$\mathbb{E}_R \text{NS}_\varepsilon(Rf|_H) \leq (1 + O(\varepsilon)) \frac{2}{\pi} d \sqrt{\varepsilon}. \quad (38)$$

Similar but simpler reasoning implies a bound on the expected average sensitivity:

Theorem 5. Let p be a polynomial of degree d and let $f(x) = \text{sgn}(p(x))$. Let $R \in \text{SO}_n$ be uniform in the Haar measure. Then

$$\mathbb{E}_R \text{AS}(Rf|_H) \leq (1 + O(1/n)) \frac{2}{\pi} d \sqrt{n}. \quad (39)$$

Proof. The angle between two adjacent corners of the hypercube is

$$r = \cos^{-1}(1 - 2/n) = \frac{2}{\sqrt{n}} (1 + O(1/n)),$$

and $\mathbb{E}_R \text{AS}(Rf|_H)$ is n times the probability that p changes sign between a uniformly random pair of points r apart. Using Lemma 5 as in the proof of Theorem 4, this probability is at most $(d/\pi)r$. \square

Remark 1. For small ε , we can also prove Theorem 4 by noting that the angle r between two points x, y on the hypercube where we have flipped each bit independently with probability ε obeys $\mathbb{E}[\cos r] = 1 - 2\varepsilon$.

Remark 2. The leading constant in (39) is better than we would obtain from Theorem 2, even after optimizing the parameter α .

Remark 3. Theorems 3 and 4 immediately generalize from polynomials to any class of functions with a bound on the number of roots lying on a great circle. If there are at most b such roots, the expected noise sensitivity of the corresponding threshold function is $O(b\sqrt{\varepsilon})$ and its expected average sensitivity is $O(b\sqrt{n})$.

Acknowledgments We are grateful to Fabrice Baudoin, Laura De Carli, Costas Efthimiou, Veit Elser, Josh Grochow, Ilia Krasikov, Ryan O’Donnell, Thomas H. Parker, Dan Rockmore, and James Stokes for helpful conversations. This work was supported by NSF grants CCF-1117426 and CCF-1219117.

References

- [1] Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma and low-weight approximators for low-degree polynomial threshold functions. *Theory of Computing*, 10(2):27–53, 2014.
- [2] Christopher Frye and Costas J. Efthimiou. Spherical harmonics in p dimensions. 05 2012. URL <http://arxiv.org/abs/1205.3548>.
- [3] Jean Gallier. Notes on spherical harmonics and linear representations of Lie groups. 2013. URL <http://www.cis.upenn.edu/~cis610/sharmonics.pdf>.
- [4] Craig Gotsman and Nathan Linial. Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994.
- [5] Prahladh Harsha, Adam Klivans, and Raghu Meka. Bounding the sensitivity of polynomial threshold functions. *Theory of Computing*, 10(1):1–26, 2014.
- [6] Nobuyuki Ikeda and Shinzo Watanabe. A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14:619–633, 1977.
- [7] Gil Kalai. Noise sensitivity and chaos in social choice theory. In Gyula O.H. Katona, Alexander Schrijver, Tamás Szőnyi, and Gábor Sági, editors, *Fete of Combinatorics and Computer Science*, volume 20 of *Bolyai Society Mathematical Studies*, pages 173–212. Springer Berlin Heidelberg, 2010.
- [8] Daniel M. Kane. The Gaussian surface area and noise sensitivity of degree- d polynomial threshold functions. *Computational Complexity*, 20(2):389–412, 2011.
- [9] Daniel M. Kane. The correct exponent for the Gotsman–Linial conjecture. *Computational Complexity*, 23(2):151–175, 2014.
- [10] S. Khot, G. Kindler, E. Mossel, and R. O’Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.
- [11] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993.
- [12] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences invariance and optimality. In *Proc. 46th FOCS*, pages 21–30, 2005.
- [13] Ryan O’Donnell. Hardness amplification within NP. *Journal of Computer and System Sciences*, 69(1):68–94, 2004. Special Issue on Computational Complexity 2002.
- [14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.
- [15] Yuval Peres. Noise stability of weighted majority. 2004. URL <http://arxiv.org/abs/math/0412377>.
- [16] J. Schmid. On the affine Bézout inequality. *Manuscripta Mathematica*, 88(1):225–232, 1995.

- [17] I. R. Shafarevich. *Basic Algebraic Geometry*. Springer-Verlag, New York, NY, 1994.
- [18] G. Szegő. *Orthogonal Polynomials*. American Mathematical Society, 1939.
- [19] N. Ja. Vilenkin. *Special Functions and the Theory of Group Representations*. American Mathematical Society, 1968.

A The expected energy in terms of the Gegenbauer and Kravchuk polynomials

For $0 \leq k, h \leq n$, let $\kappa_k(h)$ denote the Kravchuk polynomial

$$\kappa_k(h) = \sum_{j=0}^k (-1)^h \binom{k}{j} \binom{n-k}{h-j}.$$

Lemma 6. Let $f : S_{n-1} \rightarrow \mathbb{C}$ be a spherical harmonic of degree ℓ , and let $R \in \mathbf{SO}_n$ be uniform in the Haar measure. Then

$$\mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 = \frac{\|f\|_S^2}{\sqrt{d_\ell}} \frac{1}{2^n} \sum_{h=0}^n \kappa_k(h) \gamma_\ell(1 - 2h/n),$$

Proof. Given $x, y \in \mathbb{Z}_2^n$, we define $x \oplus y$ as their sum in \mathbb{Z}_2^n . In terms of the embedding $\{\pm 1/\sqrt{n}\}$ of \mathbb{Z}_2^n in the unit sphere, $(x \oplus y)_i = \sqrt{n}x_i y_i$. In that case we have $\chi_k(x \oplus y) = \chi_k(x)\chi_k(y)$. Then

$$\begin{aligned} \mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 &= \mathbb{E}_R |\langle Rf, \chi_k \rangle_H|^2 \\ &= \mathbb{E}_R \left| \mathbb{E}_{x \in \mathbb{Z}_2^n} \chi_k(x) Rf(x) \right|^2 \\ &= \mathbb{E}_R \mathbb{E}_{x, y \in \mathbb{Z}_2^n} \chi(x) \chi_k(y) Rf(x) Rf(y)^* \\ &= \mathbb{E}_{x, y \in \mathbb{Z}_2^n} \chi_k(x \oplus y) \mathbb{E}_R Rf(x) Rf(y)^*. \end{aligned}$$

Let us place the north pole η at the corner 0^n of the hypercube, i.e., at $(1, \dots, 1)/\sqrt{n}$. Since the angle between x and y is the same as between η and $x \oplus y$, we can change variables to $z = x \oplus y$ and write this expectation as

$$\mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 = \mathbb{E}_{z \in \mathbb{Z}_2^n} \chi_k(z) \mathbb{E}_R Rf(z) Rf(\eta)^*. \quad (40)$$

As in Section 5, let $\mathcal{E}_\ell^{(z)}$ denote the linear operator that evaluates a spherical harmonic of degree ℓ at the point z . Writing (40) in terms of $\mathcal{E}_\ell^{(z)}$ and using (24), we have

$$\begin{aligned} \mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 &= \mathbb{E}_{z \in \mathbb{Z}_2^n} \chi_k(z) \mathbb{E}_R \mathcal{E}_\ell^{(z)}(Rf) \mathcal{E}_\ell^{(\eta)}(Rf)^* \\ &= d_\ell \mathbb{E}_{z \in \mathbb{Z}_2^n} \chi_k(z) \mathbb{E}_R \left[\left\langle \gamma_\ell^{(z)}, Rf \right\rangle_S \left\langle Rf, \gamma_\ell^{(\eta)} \right\rangle_S \right], \end{aligned}$$

Now we apply Schur's lemma (23) and the formula (16) for the inner product of two rotated Gegenbauer polynomials to obtain

$$\begin{aligned} \mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 &= \|f\|_S^2 \mathbb{E}_{z \in \mathbb{Z}_2^n} \chi_k(z) \left\langle \gamma_\ell^{(z)}, \gamma_\ell^{(\eta)} \right\rangle_S \\ &= \frac{\|f\|_S^2}{\sqrt{d_\ell}} \mathbb{E}_{z \in \mathbb{Z}_2^n} \chi_k(z) \gamma_\ell^{(\eta)}(z). \end{aligned} \quad (41)$$

Finally, we write the expectation over z as a sum over z 's Hamming weight h . For each $0 \leq j \leq |k|$, there are $\binom{k}{j} \binom{n-k}{h-j}$ points $z \in \mathbb{Z}_2^n$ with Hamming weight h and with j ones in the support of k , in which case $\chi_k(z) = (-1)^j$. Thus

$$\sum_{\substack{z \in \mathbb{Z}_2^n \\ |z|=h}} \chi_k(z) = \sum_{j=0}^k (-1)^j \binom{k}{j} \binom{n-k}{h-j} = \kappa_k(h).$$

Finally, we have $\eta \cdot z = 1 - 2h/n$. Thus

$$\mathbb{E}_R \left| \widehat{Rf}|_H(k) \right|^2 = \frac{\|f\|_S^2}{\sqrt{d_\ell}} \frac{1}{2^n} \sum_{h=0}^n \kappa_k(h) \gamma_\ell(1 - 2h/n),$$

completing the proof. \square

B A calculus proof of Lemma 4

Proof. As in the main text, applying the Laplacian on \mathbb{R}^{n-1} to a spherically-symmetric function $K_t(u, v) = f(r)$ gives

$$\Delta_{\mathbb{R}^{n-1}} f = \frac{\partial^2 f}{\partial r^2} + (n-2) \frac{1}{r} \frac{\partial f}{\partial r}.$$

If $f(u, v) = f(r)$ is a probability distribution on v , the second moment of r is an integral over spherical shells of radius r . A shell of thickness dr has volume $\Omega_{n-2} r^{n-2} dr$, giving

$$1 = \Omega_{n-2} \int_0^\infty f(r) r^{n-2} dr$$

$$\mathbb{E}[r^2] = \Omega_{n-2} \int_0^\infty f(r) r^n dr.$$

If f obeys the differential equation

$$\frac{\partial f}{\partial t} = \Delta_{\mathbb{R}^{n-1}} f,$$

then integrating by parts gives

$$\begin{aligned} \frac{\partial}{\partial t} \mathbb{E}[r^2] &= \Omega_{n-2} \int_0^\infty \Delta_{\mathbb{R}^{n-1}} f(r) r^n dr \\ &= \Omega_{n-2} \left[\int_0^\infty \frac{\partial^2 f}{\partial r^2} r^n dr + (n-2) \int_0^\infty \frac{\partial f}{\partial r} r^{n-1} dr \right] \\ &= \Omega_{n-2} \left[n(n-1) \int_0^\infty f(r) r^{n-2} dr - (n-2)(n-1) \int_0^\infty f(r) r^{n-2} dr \right] \\ &= 2(n-1) \Omega_{n-2} \int_0^\infty f(r) r^{n-2} dr \\ &= 2(n-1). \end{aligned}$$

Integrating over t recovers the fact that, after undergoing Brownian motion for time t ,

$$\mathbb{E}[r^2] = \sum_{i=1}^{n-1} \mathbb{E}[x_i^2] = 2(n-1)t.$$

Now we carry out the same calculation for S_{n-1} . Place u at the north pole, and let r denote the geodesic distance (equivalently, the angle) between u and v . Then $K_t(u, v) = f(r)$ is a zonal function, and applying the Laplace-Beltrami operator gives

$$\Delta_{S_{n-1}} f = \frac{\partial^2 f}{\partial r^2} + (n-2) \frac{\cos r}{\sin r} \frac{\partial f}{\partial r}.$$

We again integrate over annular shells at distance r . A shell of angular thickness dr has volume $\Omega_{n-2} \sin^{n-2} r dr$, so

$$1 = \Omega_{n-2} \int_0^\pi f(r) \sin^{n-2} r dr$$

$$\mathbb{E}[r^2] = \Omega_{n-2} \int_0^\pi f(r) r^2 \sin^{n-2} r dr.$$

Applying

$$\frac{\partial f}{\partial t} = \Delta_{S_{n-1}} f,$$

a lengthier integration by parts (keeping in mind that the boundary terms are zero) yields

$$\begin{aligned} \frac{\partial}{\partial t} \mathbb{E}[r^2] &= \Omega_{n-2} \int_0^\pi \Delta_{\mathbb{R}^{n-1}} f(r) r^2 \sin^{n-2} r dr \\ &= \Omega_{n-2} \left[\int_0^\pi \frac{\partial^2 f}{\partial r^2} r^2 \sin^{n-2} r dr + (n-2) \int_0^\pi \frac{\partial f}{\partial r} r^2 \sin^{n-3} r \cos r dr \right] \\ &= 2 \Omega_{n-2} \int_0^\pi f(r) (\sin^{n-3} r) (\sin r + (n-2)r \cos r) dr \\ &\leq 2(n-1) \Omega_{n-2} \int_0^\pi f(r) \sin^{n-2} r dr \\ &= 2(n-1). \end{aligned}$$

where in the second-to-last line we used $r \cos r \leq \sin r$ for $r \in [0, \pi]$. Integrating over t then gives

$$\mathbb{E}[r^2] \leq 2(n-1)t,$$

and noting that $\mathbb{E}[r] \leq \sqrt{\mathbb{E}[r^2]}$ completes the proof. \square